Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2017-238-RC2, 2018

Geoscientific Model Development Discussions

Interactive comment on "On the importance of multiple-component evaluation of spatial patterns for optimization of earth system models – A case study using mHM v5.6 at catchment scale" *by* Julian Koch et al.

Anonymous Referee #2

Received and published: 2 January 2018

The manuscript by Koch et al. proposes a multicomponent metric for evaluation and optimization of a hydrological model which can be used for any spatial pattern comparison. The topic is of interest for GMD and the manuscript is well structured, the conclusions well supported by adequate figures. I have no major concerns about the manuscript but a couple of suggestions that may help to improve the manuscript.

The two major comments are:

1) Title: The title emphasizes that it is a method for Earth system models. While the

manuscript strongly focusses on hydrological models. I am not a hydrologist and I found the Introduction too focussed on hydrological models and not very interesting for Earth system modellers. The title suggests a stronger overall discussion of Earth system models, while the whole paper is mainly about hydrological models, in the introduction as well as in the discussion. I suggest to remove the reference to Earth system models in the title to not raise wrong expectations.

2) your manuscript does not mention data uncertainty, while this could/should be a major component of a comparison metric too. if the model is within the uncertainty of observations further optimization would be overfitting. As more and more datasets provide data uncertainties, the possibility to include this information can be a major advantage over other metrics.

Specific comments:

There are a number of grammar and spelling errors throughout the manuscript. As Copernicus offers an editing service I do not detail these errors here.

p.1 l. 20: " to the optimizer", the optimization issue was not introduced before and is not relevant here. stand-alone metrics do not only fail to provide the necessary information to optimizers, but also an evaluation or calibration can suffer from only one quantified characteristic.

p.2 l. 1-3: I don't understand, earth system models usually have 2 spatial dimensions, but I dont see why they are and obstacle for modelling efforts. Do you mean the spatial scale or resolution? Even then I am not sure whether this is the major obstacle in general. Maybe it is for hydrological models? Otherwise please add a reference. It does not get clear from this sentence why this should be the case.

p.2. l. 6-9. These developments could be interesting if you would give more detail. It would also put your work better in the context. Do these approaches already use multicomponent metrics? what are the differences between the approaches of spatial

pattern oriented model evaluation? These examples are all from the field of hydrology? No other field of research has been dealing with such metrics?

p.2 l. 9-11: Strange. In Earth system modelling spatial and temporal scales are quite related. For instance the necessary temporal time step depends on the spatial resolution. also parameterizations might require adjustments due to changes in temporal or spatial resolutions. Maybe this is very specific for hydrological models?

p.2 l. 15-16: It might depend on the application of the model, sometimes the spatial pattern might even be irrelevant and a good temporal performance is sufficient. At some later point you mention that the necessary performance depends on the application of the model, but it might be useful to mention this already earlier in the introduction.

p.3 l. 1-5: are the requirements for earth system models and hydrological models the same? you claim your studies findings are imporant for earth system models but all your requirements and testing seem very focussed on hydrological models.

p.3 l. 9-12: if your variable has different units, ok. but if the unit is the same you might want your model to have the same mean or at least not a large deviation. That would then require an additional metric? how would you merge it then with your multicomponent metric?

p.3 l. 15: the possibility to include data uncertainties could be another point. remote sensing data inlude considerable uncertainties, optimizing the model by treating the "observed data" as the truth can lead to overfitting or biased model parameters especially if the uncertainties in the data scale with another important variable or increase with increasing values of the variable.

p.4 l.30: this seems your way to partly deal with the data uncertainty.

p.5, l.17, "source of information" this seems to be the wrong expression, probably a single metric or a single characteristic? single source of information sounds to me like using only one dataset to compare the model with as opposed to using multiple

datastreams to optimize or evalute the model.

p.6, l.3: why are you doing a sensitivity analysis? Is this to select a limited set of parameters for the optimization? if yes please explain.

p.6 l. 22-25: This seems to be a result, please move this paragraph.

p. 12. l. 14: The insensitivity to bias can also be a disadvantage, in many cases the optimized model is desired to be unbiased. p.12, l. 15: if the units differ, it might depend how the two units relate to each to other. it certainly is ok if they linearly scale. How about a nonlinear relationship? How about a possible change in sign as for instance with celsius and kelvin? if the mean temperature in celsius would go towards zero you would get difficulties for the beta part of your metric?

Reproduceability: Will you provide your model outputs, observations used and analysis scripts?

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-238, 2017.